DOI:
10.1111/bjd.16265

Document Version
Peer reviewed version

Manuscript for British Journal of Dermatology

TITLE:
An Oral Disease Severity Score (ODSS) validated for use in oral pemphigus vulgaris

MANUSCRIPT WORD COUNT: 3475
TABLES: 3
FIGURES: 7

M Ormond[1], H McParland[1], ANA Donaldson[2], M Andiappan[2], RJ Cook[1], M Escudier[1, 3], E Hullah[1], J Higham[4], R McMillan[5], J Taylor[6], PJ Shirlaw[1], SJ Challacombe[1, 3], JF Setterfield[1, 3, 7].

[1] Department Oral Medicine, Dental Institute, Guy's and St Thomas NHS Foundation Trust and King's College London, London, UK
[2] Biostatistics and Research Methods Centre, Dental Institute, King's College London, UK
[3] Mucosal and Salivary Biology Division, Dental Institute, King's College London, UK
[4] Birmingham Dental Hospital, Birmingham, UK
[5] Eastman Dental Hospital, UCLH / Eastman Dental Institute, UCL, London, UK
[6] University Dental Hospital, Manchester, UK
[7] St John's Institute of Dermatology, Guy's and St Thomas's NHS Foundation Trust, London, UK

Correspondence to:
Jane Setterfield
E-mail: jane.setterfield@kcl.ac.uk
Department of Oral Medicine, Guy's Hospital, Guy's and St Thomas's NHS Foundation Trust, London, UK

No external application for funding was been made.

No conflicts of interest to declare.

What is already known about this topic?

- Current pemphigus vulgaris (PV) disease activity indices include the recently validated Pemphigus Disease Area Index (PDAI) and the Autoimmune Bullous Skin Disorder Intensity Score (ABSIS).
- The Oral Disease Severity Score (ODSS) has been demonstrated to have good inter- and intra-observer reliability in both lichen planus (LP) and mucous membrane pemphigoid (MMP).

What does this study add?

- The ODSS is shown to be a thorough, sensitive, yet quick assessment tool for oral involvement in PV.
- Its versatility for use additionally in MMP and LP is an added advantage over previously validated methods.

SUMMARY

Objectives

The primary aim of this study was to validate the Oral Disease Severity Score (ODSS) for the assessment of oral involvement in pemphigus vulgaris (PV). A secondary aim was to compare the inter – intra- observer variability and ease of use with the Physician's Global Assessment (PGA) and the oral scoring methods used in Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) and the Pemphigus Disease Area Index (PDAI).

Methods

15 patients with mild to moderately severe oral PV were scored for disease severity by 10 oral medicine clinicians using the ODSS, PGA and the oral sections of ABSIS and PDAI. Two clinicians re-scored all patients after a minimum two-hour interval.

Results

Inter-observer reliability was assessed using an intra-class correlation coefficient (ICC). For the ODSS total score the ICC was 0.83, PDAI (oral total activity) 0.79 ABSIS (oral total) 0.71 and PGA was 0.7. Intra-observer agreement between initial scoring and re-scoring of the same subject by two clinicians demonstrated an ICC for each of 0.97 and 0.96 for ODSS total score; 0.99 and 0.82 for the PDAI oral activity; 0.86 and 0.45 for the ABSIS total and 0.99 and 0.64 for the PGA. Convergent validity was good with a correlation coefficient of greater than 0.5 ($p<0.0001$). The mean time (SD) (seconds) taken to complete each scoring method was: ODSS 76±37; PDAI 117±16; ABSIS 75±19.

Conclusion

This study has validated the ODSS for the assessment of oral PV. It has shown superior inter- and intra-observer reliability to PDAI, ABSIS and PGA and is quick to perform.

INTRODUCTION

Pemphigus vulgaris (PV) is a rare autoimmune bullous disease, which can present with or develop into a condition with severe and recalcitrant oral mucosal lesions.[1] Optimal management of such cases relies upon careful clinical assessment and documentation without which there may be a delay in the recognition of therapeutic response or of treatment failure.

To determine the optimal management for PV, numerous clinical trials have been undertaken. However, assessment of efficacy has been hampered by the lack of validated clinical outcome measures particularly where mucosal sites have been affected. A Cochrane review revealed that over 116 outcome measures have been described in 96 articles and concluded that there was insufficient evidence to determine which therapy is optimal.[2] Since that time there have been further systematic reviews, which similarly have been unable to clearly establish optimal therapeutic guidelines.[1, 3] Multiple PV disease activity indices including the Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) and Pemphigus Disease Area Index (PDAI) have been validated.[4, 5, 6] While these methodologies include mucosal scores, our aim was to compare their sensitivity to a more detailed methodology for scoring oral mucosal sites. Given the recalcitrant nature of oral lesions in severe PV it is imperative that a formally validated, sensitive and user friendly scoring methodology for the oral mucosa be available to facilitate international collaboration, multicentre studies as well as providing more useful comparison of clinical outcomes in smaller case series. Thereafter, sequential scores over time relating disease activity with other outcome measures including quality of life measures and desmoglein antibody titres will be most informative.

The Oral Disease Severity Score (ODSS) is a comprehensive scoring methodology devised by the Oral Medicine group at Guy's Hospital as part of a strategy of having disease severity scores applicable to most, if not all, oral mucosal diseases. It was first developed from a scoring system devised for multi-site mucous membrane pemphigoid (MMP).[7] The ODSS records the presence of lesions and degree of activity at multiple oral sites. Additionally it includes a subjective assessment of the patient's degree of oral pain over the preceding week. Previous studies have demonstrated that the ODSS has good inter- and intra-observer reliability in both lichen planus (LP) and MMP.[8, 9] In addition it has been shown to be valuable in the assessment of therapeutic response over time in severe mucosal LP and PV.[10, 11]

The primary aim of this study was to validate the ODSS for the assessment of oral involvement in PV. We invited oral medicine clinicians from four centres in the UK with a range of experience in scoring methodologies and included patients with a range of disease severity of oral pemphigus. In the validation of the ODSS we investigated the inter- and intra-observer variability and ease of use. The ODSS was used in parallel with the PGA and the oral part of two recently proposed and validated systems for autoimmune bullous diseases, the ABSIS and PDAI. We additionally

sought to assess convergent validity between PDAI and ODSS, ABSIS and PGA .[4, 5, 6, 12]

METHODS

Research ethics approval was obtained (REC15/ES/0038). The study was performed within the Oral Medicine department at Guy's Hospital, London. Ten oral medicine clinicians from four UK Oral Medicine centres were involved. Fifteen patients were scored using the ODSS in addition to the oral sections of the ABSIS, the PDAI and PGA. Patients were scored during the course of one day. Each patient remained in one room with an assistant who recorded the scores and timed each methodology. Clinicians rotated until all patients had been scored. Two clinicians re-scored all patients after a minimum two-hour interval in order to reduce recall (see statistical methods below). Twelve sets of scores were recorded for each patient.

Physicians

Clinicians participating in the study were either Consultants in Oral Medicine (n=8) or Oral Medicine trainees (n=2). 6 clinicians were both medically and dentally qualified with one additionally a practicing dermatologist, while the remaining 4 clinicians were dentally qualified alone. 5 clinicians routinely used the ODSS in clinical practice while the remaining 5 did not. PDAI and ABSIS were not routinely used by any of the clinicians. Prior to the study day, a set of training clinical slides demonstrating the ODSS system, ABSIS and PDAI was circulated to all the clinicians. On the study day, clinicians met with the Chief Investigator for a detailed discussion of each methodology, review of the clinical slide set and to familiarise themselves with the calibration of each system. All clinicians examined all of the patients once and 2 clinicians examined all of the patients twice (with a 2 hour interval to reduce recall).

Patients

Fifteen adult patients (aged 18-80) with a confirmed diagnosis of predominantly oral mucosal PV (based on clinical findings, histopathology and positive direct immunofluorescence as previously described[13]) participated in the study. Patients were recruited consecutively from the Oral Medicine department. The visit replaced one of their routine follow-up appointments. Thirteen had mild to moderately severe oral disease; one was in clinical remission and one had severe oral disease. All were on systemic treatment (9 mycophenolate mofetil, 4 azathioprine and 1 rituximab with or without prednisolone) at the time of the study.

Oral Disease Severity Score (ODSS)

The ODSS is a comprehensive oral scoring system previously validated for lichen planus (Fig. 1).[8] It has additionally been used sequentially in PV and LP. [10, 11] It divides the mouth into 17 sites weighted according to area of possible involvement and allocated a site score of 0-2. The sites include the outer / inner lips, left and right buccal mucosa 6 gingival segments, hard palate (left / right or both), soft palate (left /

right or both), dorsum tongue (left / right or both), left ventral tongue, right ventral tongue, floor of mouth (left / right or both) and oropharynx. Individual sites (or units of a site) are then allocated an activity score (0-3), reflecting mild inflammation (minimal erythema or a white 'healing' mucosa) = 1 (Fig. 2a); moderate inflammation (marked erythema but no ulceration) = 2 (Fig. 2b) and ulceration = 3 (Fig. 2c). Additionally a subjective assessment of the patient's oral pain in the preceding week is included (verbal rating scale of 0-10). The theoretical maximum total score is 106; however greater than 95% of patients would be expected to have scores in the range from 0 to 60 representing a clinical range from remission to severe disease.

Autoimmune Bullous Skin Disorder Intensity Score (ABSIS)

The ABSIS is designed to assess the extent and quality of skin and oral lesions (Fig. 3). It has a total score of 206 (150 points for skin involvement, 11 for oral and 45 for subjective discomfort). For the present study we used only the oral mucosal part of the system. This evaluates 11 distinct anatomical sites. The presence or absence of any lesion (blister / erosion) is allocated a score of 0 or 1 with a total score of 11. The second part of the score is for severity of symptoms which details the discomfort whilst eating and drinking maximum score 45. A higher score represents more severe disease.[4]

Pemphigus Disease Area Index (PDAI)

The PDAI was developed by the International Pemphigus Committee to capture the spectrum of PV (Fig. 3).[5, 6] The skin, scalp and mucosa are scored separately for both the number of lesions (ulcers or erosions) present and evidence of damage (skin and scalp). There is no score for erythema / damage in the oral mucosa. The total possible score is 250 (120 for skin, 10 for scalp and 120 for mucosal site activity).[5] For the purpose of this study only oral mucosal surfaces were scored with a possible total of 90 (excluding the eyes, nose and anogenital areas).

Physician's Global Assessment (PGA)

The Physician Global Assessment is a simple scoring method for inflammatory skin disease (Fig. 5). It is a ten-point analogue scale from 0 (perfect health) to 10 (worst condition imaginable). The clinicians score the participants on a global impression of their disease. This score method has been validated in a number of diseases including psoriasis and eczema.[6, 12, 14]

Time for scoring methodology completion

An independent assistant scored the time (seconds) taken by each clinician to obtain a disease severity score for each method using a stopwatch.

Statistical Methods

Inter-observer reliability was undertaken with 10 observers (clinicians) scoring all patients with each of the four scoring methodologies. A sample size of 15 subjects was required to achieve intra-class correlations (ICC) of 0.77 for the inter-observer reliability.

Intra-observer reliability was tested with 2 replications per subject (as per test-retest) with a minimum of 2 hours between scores to minimise the risk of recall. Since the involvement was more onerous (burden, time or money resources, etc.) for the rater than for the subject, taking rater as fixed in the factorial design was more efficient. We fixed the number of raters to 2 and found the sample size required in terms of the number of subjects (which are assumed to be a random sample from the population of subjects). With both raters performing 2 replications in each methodology in all the subjects, a total of 15 subjects provides 80% power to detect an ICC difference of 0.50 (relative to a null value of 0.20). Anticipating an ICC of 0.85, 15 subjects with 2 replications will yield a width of 0.30 in the 95% confidence interval.

Multilevel models were used to quantify intra- and inter-observer reliabilities of the continuous measures. Assessment for the level of agreement in terms of the intra-class correlation coefficients for ordinal or continuous measures followed well established benchmark limits (Fleiss and Altman's benchmark scales).[15, 16] Landis-Koch's benchmark values were followed when Kappa coefficients were used for categorical outcomes. In all cases, for more rigour, in addition to the point estimate, we took into account the lower bound of the 95% confidence interval.[17] Convergent validity was calculated using the Spearman rank correlation coefficient.

RESULTS

Fifteen patients (f11:m4) with confirmed PV were included. Their mean (SD) age was 56±14.2 years (range 23-77).

The distribution of scores:

ODSS:

The mean (SD) total ODSS score was 22.3±12.8, (range 0-68); median [IQR] 22 [12.75-29], reflecting mild to moderately severe disease in 13/15. In one patient no lesions were identified (score 0), while another with very severe disease had a score of 68.

PDAI:

The mean PDAI activity score for oral mucosa was 14.6±7.7, (range 0-42); median 14.6 [11.0-16.5].

ABSIS:

The mean total score was 12.5±1.1, (range 0-43); median 11.7 [5.4-17.5]. The mean score for oral involvement; was 3.6 ±2.3; median 4 [2-5]. For disease severity the mean was 8.4±7.0; median 8 [0.5-12.5].

PGA:

The mean score was 6.1±0.9, (range 0-9); median 3 [1-8].

Reliability

Test-retest reliability (Table 1)

Intra-observer agreement between initial scoring and re-scoring of the same subjects demonstrated an intra-class correlation coefficient (ICC) for each of the two examiners of 0.96 and 0.97 for ODSS total, 0.95 and 0.97 ODSS site, 0.95 and 0.98 ODSS activity and 0.9 and 0.97 for ODSS pain. As only oral mucosal sites were

scored for the PDAI, the total PDAI reflects activity only with no damage score. The PDAI ICC was 0.82 and 0.99. The ABSIS total ICC was 0.45 and 0.72 with ABSIS involvement 0.9 and 0.94 and ABSIS severity 0.44 and 0.91. The ICC for PGA was 0.64 and 0.99.

Inter-observer reliability (Table 2)

The ICC for the total ODSS score was 0.83 (0.71-0.94); ODSS site 0.69 (0.52-0.86), ODSS activity 0.83 (0.72-0.94), and ODSS pain 0.9 (0.84-0.97). The PDAI activity ICC was 0.79 (0.65-0.92). For ABSIS the ICC for total score was 0.71 (0.55-0.88); oral involvement 0.72 (0.57-0.83) and severity 0.67 (0.5-0.85). The PGA ICC was 0.7 (0.54-0.87).

Convergent Validity (Tables 3)

The convergent validity between the PDAI activity (gold standard) and the other indices are detailed in Table 3. There was good convergent validity for all indices (ODSS total 0.70, ABSIS total 0.51 and PGA 0.77, $p<0.0001$).

Time taken for completion of disease scoring methodologies

The mean (SD) time to obtain a disease severity score for ODSS was 76 ±37; PDAI 117±16 and ABSIS 75±19 seconds.

DISCUSSION

This study has shown that the ODSS is a valid method for assessment of disease severity for oral PV and has a higher inter- and intra-observer reliability than the previously validated methodologies PDAI, ABSIS and PGA.

The patient sample used in this study reflects the reported sex and age distribution of PV with a broad range of oral disease severity (mild to severe) despite all being on systemic treatment.[18] The mean and range of scores for ABSIS and PDAI were detailed by Rosenbach in their validation study and reflect a similar and potentially milder group of patients. [6]

In terms of reliability of the methodologies, the intra-observer scores were excellent, almost perfect (ICC>0.9) for all parameters for ODSS (0.95-0.97) for total score, PDAI (0.82-0.99) and for ABSIS total score (0.86-0.45). Intra-observer scores were given a benchmark value of good / substantial for PDAI activity and fair / moderate for total score for ABSIS. For PGA the benchmark value was good / substantial. Rosenbach's data showed that for intra-observer reliability, using the test-retest method (2 replications), the ICC was 0.98 (0.97-1.0) for PDAI mucous membrane activity and for ABSIS oral involvement was 0.99 (0.97-1.0). ABSIS subjective discomfort was not calculable due to a lack of variability among subjects. Our data has shown better intra-observer scores for ODSS than the other methods examined.

For inter-observer reliabilities the ODSS total score was ICC 0.83. Benchmarking values were very good or excellent for total score, activity and pain. For PDAI

activity the ICC was 0.79 (very good); for ABSIS total score ICC was 0.71 (good / substantial) and for PGA 0.7 (good / substantial). Rosenbach reported an ICC of 0.84 for PDAI; 0.85 for extent of mucosal disease with ABSIS, and 0.89 for subjective involvement. In our study the ODSS was more reliable among the scoring clinicians than the three other methods.

The average time taken for clinicians to complete each of the scoring tools was less than 2 minutes; the longest completion time being PDAI. The time difference for completion of both the PDAI and ABSIS may in part be due to a lack of familiarity with the scoring systems as the clinicians did not routinely use either system nevertheless, the total time for all three was considered acceptable in a routine clinic. Furthermore, there was no significant difference in the mean time taken by clinicians scoring the ODSS clinically for the first time compared with those familiar with the system thus lack of familiarity should not have significantly influenced these data.

In terms of convergent validity using PDAI as the previous 'gold standard', ODSS total score demonstrated good correlation with the PDAI activity score (correlation coefficient 0.7 $p < 0.0001$).

There are some limitations to the study. A valid outcome measure for PV should ideally be reliable, discriminatory, sensitive, accurate, feasible, close to a gold standard and have excellent external validity.[19, 20] Potentially scores at the lower end of the spectrum for all methodologies might reduce the sensitivity and reproducibility of each scoring system. However, the fact that ODSS details more oral sites than either ABSIS or PDAI, has a higher maximum score and combines both a semi-quantitative score (site) with a qualitative score (activity) permits detailed differentiation between patients. Ideally the cohort would have included untreated severe patients but this would have necessitated withholding treatment. Secondly, the scoring instruments compared here looked only at the oral mucosal lesions and therefore would need to be undertaken alongside a skin and other mucosal site score. Thirdly intra-observer reliability was tested with a minimum 2-hour interval. This interval may allow recall bias and ideally this would be longer but more than a day or two might be associated with changes in disease activity. Furthermore patients would need to reattend which adds an impractical level of complexity to the study.

In terms of external validity, the sample of patients included would ideally be randomly selected from a large cohort. Our sample was selected consecutively from those under current follow-up, with predominantly active disease and who were able to attend on the study day. We did not exclude any patient over and above those constraints. Clinicians with an interest in the field of immunobullous disease were invited and again were not randomly selected. The intra-observer reliability may have been improved by increasing the replications including those familiar and unfamiliar with the scoring methodologies. However using experienced clinicians in scoring methodologies is likely to have had a positive effect on all indices. Furthermore, this

provided more data as there were a further 30 scores to analyse than just a further 10 if each observer had just rescored one patient. Finally, sequential use of ODSS in case reports and a small case series has demonstrated efficacy of treatment in recalcitrant PV and LP. It now needs to be tested sequentially alongside PDAI and ABSIS to demonstrate sensitivity to change in disease activity and response to treatment in this cohort. [10, 11, 21]

We asked clinicians to comment on their experiences of each method. They all reported the ODSS to be the easiest method to use and felt that it more accurately and objectively recorded the extent of oral disease in PV. All would consider incorporating it into their routine assessment. The PDAI was considered to have potential for variable interpretation of the clinical features. The necessity of a lesion to be a blister or erosion precluded those lesions that were almost healed. It was also difficult to know how to score the number of lesions if confluent on the gingiva and including a few teeth amounting to <2cm, or if patchy and multiple. The ABSIS score was much easier to reflect severity, though it also required lesions to be blisters or erosions, thereby losing some sensitivity in white / healing areas. It had a substantial subjective component requiring the patient to report symptoms with foods and the reply sometimes depended upon how the questions were put to the patient. The system seemed weighted too strongly on this subjective component, which was not felt to allow for accurate recording of clinical severity. Finally the PGA, while simple, was felt to offer little information regarding the objective oral involvement of PV and potentially would be less valuable for sequential monitoring of disease.

The ODSS has been used in a sequential study of 23 patients and demonstrated a positive association between oral disease activity and salivary antibodies to Dsg3.[21] ODSS has also previously been used to demonstrate cumulative efficacy to rituximab in patient with recalcitrant oral pemphigus where serial serum IgG Dsg3 titres followed clinical scores over several years follow up.[11] In a cross-sectional study of PV patients comparing the inter-rater validity of PDAI, the Pemphigus Vulgaris Activity Score and ABSIS together with convergent validity according to anti-Dsg values, both inter-rater reliability and convergent validity were highest with PDAI.[22] In analysis of sequential serum samples from PV patients, the anti-Dsg3 indices showed a correlation with PDAI scores,[23] while in a further study no correlation was found with either PDAI or ABSIS.[24] Among predominantly mucosal PV patients, there was no correlation between either serum or salivary anti-Dsg1 or 3 antibodies and the total objective component of ABSIS, however serum anti-Dsg1 did correlate with cutaneous ABSIS.[25]

As PV is a rare disease, establishing optimal therapeutic regimens has been very difficult. Multicentre collaborative trials are therefore essential.[20] However; the lack of a universally accepted outcome measure has been highlighted in a recent systematic review.[1] The PDAI is considered to be the optimal method for multisite disease in use to date. While patients with PV may initially have both mucosal and

cutaneous lesions, many are left with pure oral disease, which is often severe and debilitating. We propose that a more sensitive method of assessing oral PV needs to be in place for these cases. The ODSS was first developed and published for use in LP and its use extended as part of a strategy of developing disease severity scores for a number of oral mucosal diseases.[8, 28, 29, 30] It incorporates both objective measures of disease activity and severity as well as including patient subjective data allowing for a comprehensive appraisal of mucosal disease.[8] The ODSS has been used in our department for more than 10 years and has been externally evaluated for use in MMP and has shown efficacy of treatment in PV and LP. [9, 10, 11]

In this study ODSS has been shown to be a reliable and reproducible tool for recording oral PV disease activity with inter- and intra-rater reliability at least as good as PDAI for oral lesions. It is simple to use and facilitates detailed recording of both the site and the severity of lesions including those that are healing. We propose that this scoring tool would be a useful supplement for future multicentre studies as well as recording sequential disease activity in the clinic. We are currently evaluating data on sequential scores using all 4 methods over a 1 year follow up to clarify which is the most sensitive to subtle changes in disease activity.

REFERENCES

1. McMillan R, Taylor J, Shephard M *et al*. World Workshop on Oral Medicine VI: a systematic review of the treatment of mucocutaneous pemphigus vulgaris. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2015;120(2):161-71.e20.

2. Martin L, Murrell DF. Measuring the immeasurable: a systematic review of outcome measures in pemphigus. *Paper Australas J Dermatol*. 2006;47(Suppl1):A32–3.

3. Ni Riordain R, Shirlaw P, Alajbeg I *et al*. World Workshop on Oral Medicine VI: Patient-reported outcome measures and oral mucosal disease: current status and future direction. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2015;120(2):161-71.e20.

4. Pfütze M, Niedermeier A, Hertl M, Eming R. Introducing a novel Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) in pemphigus. *Eur J Dermatol*. 2007;17(1):4-11.

5. Murrell DF, Dick S, Ahmed AR *et al*. Consensus statement on definitions of disease, end points and therapeutic response for pemphigus. *J Am Acad Dermatol*. 2008;58:1043-6.

6. Rosenbach M, Murrell D, Bystryn JC *et al*. Reliability and Convergent Validity of Two Outcome Instruments for Pemphigus. *J Invest Dermatol*. 2009;129:2404-2410.

7. Setterfield J, Shirlaw PJ, Kerr-Muir M *et al.* Mucous membrane pemphigoid: a dual circulating antibody response with IgG and IgA signifies a more severe and persistent disease. *Br J Dermatol*. 1998;138:602-610.

8. Escudier M, Ahmed N, Shirlaw P *et al*. A scoring system for mucosal disease severity with special reference to oral lichen planus. *Br J Dermatol*. 2007;157(4):765-770.

9. Reeves GMB, Lloyd M, Rajlawat BP *et al*. Ocular and oral grading of mucous membrane pemphigoid. *Graefes Arch Clin Exp Ophthalmol*. 2012;250:611-618.

10. Wee J, Shirlaw PJ, Challacombe SJ, Setterfield JF. Efficacy of mycophenolate mofetil in severe mucocutaneous lichen planus: a retrospective review of 10 patients. *Br J Dermatol*. 2012;167:36-43.

11. Greenblatt DT, Benton EC, Groves RW, Setterfield JF. Crescendo response to Rituximab in oral pemphigus vulgaris: a case with 7-year follow-up. *Clin Exp Dermatol*. 2016 Jul;41(5):529-32.

12. Langley RG, Ellis CN. Evaluating psoriasis with psoriasis area and severity index, psoriasis global assessment, and lattice system physician's global assessment. *J Am Acad Dermatol*. 2004;51:563–9.

13. Bhogal B, Wojnarowska F, Black MM *et al.* The distribution of immunoglobulins and the C3 component of complement in multiple biopsies from the uninvolved and perilesional skin in pemphigus. *Clin Exp Dermatol*. 1986;11:49-53.

14. Guzzo CA, Weiss JS, Mogavero HS *et al*. A review of two controlled multicenter trials comparing 0.05% halobetasol propionate ointment to its vehicle in the treatment of chronic eczematous dermatoses. *J Am Acad Dermatol*. 1991;25:1179–83.

15. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-8.

16. Altman DG. *Practical statistics for medical research*. Chapman and Hall: London, 1991.

17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.

18. Kneisel A, Hertl M. Autoimmune bullous skin diseases. Part 1: clinical manifestations. *J Dtsch Dermatol G*es. 2011;9:844-856:quiz 857.

19. Bastuji-Garin S, Sbidian E. How to validate outcome instruments for pemphigus. *J Invest Dermatol*. 2009;129:2328-2330.

20. Murrell DF, Amagai M, Weerth VP. Scoring systems for blistering diseases in practice: why bother, and which one should you use? *JAMA Derm*. 2014;150:245-247.
21. Ali S, Kelly C, Challacombe SJ, Donaldson ANA, Bhogal BS, Setterfield J. Serum and saliva IgG and IgA antibodies to Dsg3 and their relation to disease severity in mucosal pemphigus vulgaris. *Br J Dermol*. 2016;175(1):113-21.

22. Rahbar Z, Daneshpazhooh M, Mirshams Shahshahani M *et al*. Pemphigus disease activity measurements: pemphigus disease area index, autoimmune bullous skin disorder intensity score, and pemphigus vulgaris activity score. *JAMA Dermatol*. 2014;150(3):266-72.

23. Kamiya k, Aoyama Y, Shirafui Y et al. A higher correlation of the antibody activities against the calcium-dependent epitopes of desmoglein 3 quantified by ethylenediaminetetracetic acid-treated enzyme-linked immunodorbent assay with clinical disease activities of pemphigus vulgaris. *J Dermatol Sci*. 2013;70(3):190-195.

24. Patsatsi A, Kyriakou A, Giannakou A, Pavlitous-Tsiontsi A *et al*. Clinical significance of anti-desmoglein -1 and -3 circulating autoantibodies in pemphigus patients measured by Area Index and Intensity Score. *Acta Derm Venereol*. 2014;94(2):203-6.

25. De D, Khuller G, Handa S, Joshi N, Saikia B, Minz RW. Correlation between salivary and serum anti-desmoglein 1 and 3 antibody titres using ELISA and between anti-desmoglein levels and disease severity in pemphigus vulgaris. *Clin Exp Dermatol*. 2017;42(6):648-50.

26. Tappuni AR, Kovacevic T, Shirlaw PJ, Challacombe SJ. Clinical assessment of disease severity in recurrent aphthous stomatitis. *J Oral Pathol Med*. 2013;42(8):635-41.

27. Osailan SM, Pramanik R, Shirlaw P, Proctor GB, Challacombe SJ. Clinical assessment of oral dryness: development of a scoring system related to salivary flow and mucosal wetness. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2012;114(5):597-603

28. Campbell HE, Escudier MP, Patel P, Challacombe SJ, Sanderson JD, Lomer MC. Cinnamon- and benzoate-free diet as a primary treatment for orofacial granulomatosis. *Aliment Pharmacol Ther*. 2011;34(7):687-701

Table 1. Within observer (intra-observer) reliability data for each scoring methodology. *Assessment for the level of agreement in terms of the intra-class correlation coefficients followed Fleiss and Altman's benchmark scales.[16, 17] ICC= intra-class correlations

	Observer 1 ICC (95% CI)	**Observer 2 ICC (95% CI)**	**P-value**	**Overall benchmark values***
ODSS Site	0.95 (0.86-1.00)	0.97 (0.94-1.00)	0.36	Excellent
ODSS Activity	0.98 (0.95-1.00)	0.95 (0.91-1.00)	0.61	Excellent
ODSS Pain	0.97 (0.97-1.00)	0.90 (0.80-0.99)	0.43	Excellent
ODSS Total	0.97 (0.95-1.00)	0.96 (0.92-1.00)	0.53	Excellent
ABSIS Involvement	0.94 (0.89-1.00)	0.90 (0.80-1.00)	0.83	Excellent
ABSIS Severity	0.91 (0.82-1.00)	0.44 (0.03-0.86)	0.16	Fair / moderate
ABSIS Total	0.86 (0.72-0.99)	0.45 (0.04-0.86)	0.46	Fair / moderate
PDAI Activity	0.99 (0.99-1.00)	0.82 (0.65-0.99)	0.82	Good / substantial
PGA	0.99 (0.97-1.00)	0.64 (0.34-0.95)	0.80	Good / substantial

Table 2. Scores and Inter-observer reliability for each of four disease severity scoring systems, and individual components of ODSS. *Assessment for the level of agreement in terms of the intra-class correlation coefficients followed Fleiss and Altman's benchmark scales.[16, 17]
ICC= intra-class correlations
IQR= interquartile range

	Range	**Median (IQR) Mean±SD**	**Inter-observer ICC (95% CI)**	**Overall benchmark values***
ODSS Site	0-16	7 (4-10) 7±3.84	0.69 (0.52-0.86)	Good/Substantial
ODSS Activity	0-48	11 (6-17) 12.63±9.58	0.83 (0.72-0.94)	Very good
ODSS Pain	0-6	3 (2-4) 2.78±1.37	0.9 (0.84-0.97)	Excellent
ODSS Total (0-106)	0-68	22 (12.75-29) 22.3±12.8	0.83 (0.71-0.94)	Very good
ABSIS Involvement	0-10	4 (2-5) 3.65±2.3	0.72 (0.57-0.83)	Good/Substantial
ABSIS Severity	0-29.5	8 (0.5-12.5) 8.41±7.01	0.67 (0.50-0.85)	Moderate
ABSIS Total (0-56)	0-43	11.75 (5.4-17.5) 12.5±1.1	0.71 (0.55-0.88)	Good/Substantial
PDAI Activity (0-90)	0-42	13 (11.0-16.5) 14.6±7.7	0.79 (0.65-0.92)	Very good
PGA (0-10)	0-9	3 (1-8) 6.1±0.9	0.70 (0.54-0.87)	Good/Substantial

Table 3. Disease Severity scoring systems: convergent validity of ODSS, ABSIS and PGA with PDAI Activity (n=15). *Spearman rank correlation coefficient

Scoring method	Correlation Coefficient*	P-value
ODSS Total	0.70	<0.0001
ABSIS Total	0.51	<0.0001
PGA	0.77	<0.0001

Figure 1. Guy's Oral Disease Severity Score (ODSS)

Figure 2a. Hard palate showing bilateral whitening of the mucosa indicating mild activity (site score= 2, activity 1+1=2)

Figure 2b. Right buccal mucosa showing <50% affected with erythema and whitening but no frank ulceration (site score= 1, activity =2)

Figure 2c. Right buccal mucosa demonstrating areas of ulceration affecting <50% surface area (site score=1, activity=3)

Figure 3. Autoimmune Bullous Skin Disorder Intensity Score (ABSIS)

Figure 4. Pemphigus Disease Area Index (PDAI)

Figure 5. Physician's Global Assessment (PGA)

Figure 1. Guy's Oral Disease Severity Score (ODSS)

Site	Site Score	Activity Score/unit of site (0-3)
Outer lips (1)		
Inner lips (1)		
R Buccal mucosa (1 or 2)		
L Buccal mucosa (1 or 2)		
Gingivae (1 each segment)		
Lower R		
Lower central		
Lower L		
Upper R		
Upper central		
Upper L		
Dorsum of tongue (1 or 2)		
R Ventral tongue (1)		
L Ventral tongue (1)		
Floor of mouth (1 or 2)		
Hard palate (1 or 2)		
Soft palate (1 or 2)		
Oropharynx (1 or 2)		
Total		

Total Score

Total Score = Site Score + Activity Score + Pain Score (1-10) (Maximum 106)

Site Score

0 if no lesion
For the buccal mucosa - 1 if less than 50% of area affected; 2 if greater than 50% of area affected

For the dorsum of tongue, floor of mouth, hard or soft palate or oropharynx
1 unilateral
2 bilateral

Activity Score

1 mild erythema (gingivae, papillae only or <3mm along margins)
2 marked erythema (full thickness gingivitis, extensive atrophy)
3 ulceration at this site

Pain Score

Analogue scale from 0 (no discomfort) to 10 (the most severe pain they have encountered with this condition so far)
The patient is asked to provide a score reflecting their pain/discomfort as an average of the preceding week

Figure 3. Autoimmune Bullous Skin Disorder Intensity Score (ABSIS)

Oral Involvement

Extent (enter 1 for presence of lesions, 0 absence of any lesion)

Upper gingival mucosa	
Lower gingival mucosa	
Upper lip mucosa	
Lower lip mucosa	
Left buccal mucosa	
Right buccal mucosa	
Tongue	
Floor of mouth	
Hard palate	
Soft palate	
Pharynx	

(Total score ranges from 0-11)

Severity

Food	**Level**	**Factor of Discomfort** Legend for factor of discomfort 1 Pain/bleeding occurred always 0.5 Pain/bleeding sometimes occurred 0 Never experienced problems	**Severity Score**
Water	1		
Soup	2		
Yogurt	3		
Custard	4		
Mashed potatoes / scrambled egg	5		
Baked fish	6		
White bread	7		
Apple / raw carrot	8		
Fried steak / wholegrain bread	9		

(Severity score = Level multiplied by the factor of discomfort = 0-45 points)

Figure 4. Pemphigus Disease Area Index (PDAI)

PDAI – Mucous membranes

Anatomical Location	**Activity** Erosion/Blisters 0 Absent 1 1 lesion 2 2-3 lesions 5 >3 lesions or 2 lesions >2cm 10 At least 1 lesion >6cm	**Number of lesions if ≤3**
Eyes		
Nose		
Buccal mucosa		
Hard palate		
Soft palate		
Upper gingivae		
Lower gingivae		
Tongue		
Floor of mouth		
Labial mucosa		
Posterior oropharynx		
Anogenital		
Total mucosa	**/120**	

Figure 5. Physician's Global Assessment (PGA)

Please estimate the patient's mucosa related health on a 0–10 scale as you see it today

0	1	2	3	4	5	6	7	8	9	10
Perfect health										**Worst condition**